Statistics Canada | Statistique Canada

Methodology Branch

Census & Household Survey
Methods Division

Direction de la méthodologie

Division des méthodes de recensement
et d'enquêtes ménages

Canada

S-176E
c.2

EVALUATION OF COMPOSITE ESTIMATION FOR THE CANADIAN LABOUR FORCE SURVEY

S. Kumar and H. Lee

Census and Household Survey Methods Division

Methodology Branch

CHSM 85-067E

EVALUATION OF COMPOSITE ESTIMATION FOR THE CANADIAN LABOUR FORCE SURVEY[1]

S. Kumar and H. Lee[2]

This study considers the suitability of composite estimation techniques for the Canadian Labour Force Survey. The performance of a class of AK composite estimators introduced initially by Gurney and Daly is investigated for several characteristics. While the ordinary composite estimate has a large bias, the AK composite estimate is capable of reducing the bias. Composite estimates having minimum variance and minimum mean square error are compared.

1. INTRODUCTION

The Canadian Labour Force Survey (LFS) is conducted each month by Statistics Canada and is designed to produce estimates for various labour force characteristics. The LFS sample design follows a rotation scheme that permits the replacement of one-sixth of the households in the sample each month (see [7]). The sample is composed of six panels or rotation groups. A panel remains in the sample for a period of six consecutive months.

As pointed out in Bailar [1], one of the major drawbacks of composite estimation currently in use for the U.S. Current Population Survey (CPS) is its bias as compared to the simple ratio estimator for estimates of level. This bias stems from rotation group differences: the phenomenon that estimates based on data from different panels relating to the same time period do not have the same expected value. This phenomenon, often referred to as the rotation group bias, has been studied for LFS (see [2] and [6]). Recently, Huang and Ernst [4] have reported results in the context of the CPS on the performance of AK composite estimator introduced initially by Gurney and Daly [3]. A and K are

---

[1] Presented at the Joint Statistical Meetings of the American Statistical Association, the Biometric Society, the Institute of Mathematical Statistics and the Statistical Society of Canada in Toronto, August 1983.

[2] S. Kumar and H. Lee, Census and Household Survey Methods Division, Statistics Canada.

constants in the equation defining the composite estimator. Their results show improvement over the composite estimates currently in use for CPS as regards variance and bias.

The objective of this investigation is to study the suitability of composite estimation techniques for LFS. In this study the performance of different composite estimators of level and change will be investigated for the following five characteristics; in labour force, employed, employed agriculture, employed non-agriculture, and unemployed. These composite estimators are compared with the simple ratio estimator which is presently in use for LFS. The study is based on the province of Ontario data for 1980-81.

2. DEFINITIONS AND NOTATION

We are interested in estimating Y_m the number of persons in the population with a certain characteristic for the month m. Let

$y_{m,i}$ = A simple ratio estimator of Y_m based on the i-th panel ($i = 1,2,\ldots,6$). Here the i-th panel refers to the sub-sample (rotation group) that is in the sample for the i-th time. It will be referred to as the i-th panel estimator.

$d_{m,m-1}$ = estimator of change $(Y_m - Y_{m-1})$ from the month $(m - 1)$ to the month m based on five panels that are common to the months m and $(m - 1)$

$$= \sum_{j=2}^{6} (y_{m,j} - y_{m-1,j-1})/5. \qquad (2.1)$$

y'_m = AK composite estimator of Y_m defined as

$$y'_m = (1 - K + A)y_{m,1}/6 + (1 - K - \frac{A}{5}) \sum_{j=2}^{6} y_{m,j}/6$$

$$+ K(y'_{m-1} + d_{m,m-1}) \qquad (2.2)$$

where K and A are constants, and $0 \leq K < 1$.

The equation (2.2) defines a class of estimators referred to as AK composite estimators. The estimators obtained by taking A = 0 in (2.2) are referred to as K composite estimators. The simple ratio estimator, to be denoted by $\bar{y}_m$, the mean of six panel estimators can be obtained by taking A = 0 and K = 0 in (2.2). We investigate the relative performance of the optimal (minimum variance or minimum mean square error) AK composite, K composite and simple ratio estimators.

We assume the rotation group bias $E(y_{m,i}) - Y_m$ is independent of m and is a function of i. We denote this bias by α_i. Formally

$$\alpha_i = E(y_{m,i}) - Y_m. \qquad (2.3)$$

The expression for the bias of the composite estimator is given in Appendix I.

3. ASSUMPTIONS

The rotation system in the LFS is schematically described in Table 1, where the current (month m) panel i (= 1,2,...,6, denoting interview month no.) is the same as panel i - j in month m - j, provided i - j lies between 1 and 5. The immediate predecessor to panel i of month m as of month m - j is given by (6 + i - j) provided (6 + i - j) lies between 1 and 6. Likewise, the second predecessor to panel i as of month m - j is given by (12 + i - j) provided (12 + i - j) lies between 1 and 6. In general, the r-th predecessor to panel i of month m is given by (i - j + 6r) in month m - j. Note that the 0-th predecessor to a panel means the same panel in earlier months.

The expression for the variance of y'_m, i.e. $V(y'_m)$ involves the variances and covariances of various panel estimators (see Appendix II). The following variance-covariance structure for various panel estimators is assumed. The

assumptions conform to the LFS rotation pattern, illustrated in Table 1.

(i) $V(y_{m,i}) = \sigma^2$ for all m and i = 1,2,...,6,

(ii) $Cov(y_{m,i}, y_{m-j,i-j+6r}) = \gamma_j^{(r)}\sigma^2$, where i = 1,2,...,6, j > 0 and r ≥ 0, such that $6 \geq i - j + 6r \geq 1$. Here r denotes the number of predecessors to the current panel.

For r = 0, i.e., $6 > i - j \geq 1$, let $\gamma_j^{(r)} = \rho_j$ (based on overlapping panels of months m and m - j).

For r = 1, i.e., $6 \geq i - j + 6 \geq 1$, let $\gamma_j^{(r)} = \gamma_j$ (based on the current panel and its immediate predecessor j months back).

For r ≥ 2, i.e., $6 \geq i - j + 6r \geq 1$, let $\gamma_j^{(r)} = 0$ (based on the current panel and its r-th predecessor j months back).

(iii) Of interest to the development of the variance of the composite estimator y'_m are the correlation coefficients ρ_j and γ_j, both of which are assumed to be stationary; i.e. they are functions of j and not of m. It is reasonable to assume that both ρ_j's and γ_j's are positive since ρ_j's are based on characteristics of largely common households while γ_j's are based on the characteristics of households in the current month and those of their near (in many cases next door) neighbours j months back (apart from cluster rotation).

(iv) The expression for $V(y'_m)$ contains covariance terms not included in the assumptions (ii) and (iii). Some of these are:

$Cov(y_{m,i}, y_{m,j})$ for $i \neq j$, $Cov(y_{m,i}, y_{m-1,j})$ for i = 1, $j \neq 6$, and $i \neq 1$, $j \neq i - 1$, and $Cov(y_{m,i}, y_{m-g,j})$ for $g \geq 12$. These and all other covariances not defined above, including those with $\gamma_j^{(2)}$ and existing in the expression for $V(y'_m)$ are assumed to be zero.

Following these assumptions, a variance expression for the AK composite estimator was derived in terms of the above parameters. The mathematical details for derivation of the expression for the bias and variance of y'_m, and the variance of $y'_m - y'_{m-1}$ are given in the appendices.

4. RESULTS AND DISCUSSION

The quantities σ^2, ρ_j and γ_j in the expresssion for $V(y'_m)$ were replaced by their estimates (For details of the methodology for estimating ρ's and γ's, see [5]). Note that, in the Canadian LFS ρ_j's do not exist for $j \geq 6$ because of no overlapping panels. Nor do γ_j's exist for $j \geq 12$ because for $j \geq 12$, there exist 2nd or higher order predecessors to the current panel and the correlation may be taken as 0 in the developments. Estimates of ρ_j, $\hat{\rho}_j$, are given in Table 2. The estimate of ρ_5 has been obtained by extrapolating other ρ_j's as it was not possible to estimate it directly from the sample. Note that $\hat{\rho}_j$ $(j = 1,2,\ldots,5)$ is a decreasing function of j for all the five characteristics. This is consistent with what we expect intuitively about the behaviour of ρ_j's. Also ρ_j's are high for all the characteristics except "unemployed".

Table 3 gives the estimates $\hat{\gamma}_j$ of γ_j. The estimates $\hat{\gamma}_5$ and $\hat{\gamma}_{11}$ were obtained respectively by interpolating and extrapolating other $\hat{\gamma}_j$'s. Intuitively, we expect γ_j's to decrease with j for each characteristic. We observe that this is not the case with $\hat{\gamma}_j$'s. Although $\hat{\gamma}_j$'s do not exhibit monotonic decreasing behaviour, we point out that whenever the difference $(\hat{\gamma}_{j+1} - \hat{\gamma}_j)$ is positive, its magnitude is very small. The positiveness of these differences could be due to the sampling variability rather than a real positiveness of $(\gamma_{j+1} - \gamma_j)$.

In the following discussion, the term relative efficiency of AK composite (or K composite) estimator refers to its efficiency relative to the simple ratio estimator.

Tables 4A and 4B give the results of comparing the estimated variances of three estimators. These are: (i) optimal AK composite estimator, i.e., an estimator having minimum variance among the class of estimators defined by (2.2), (ii) optimal K composite estimator (obtained by taking A = 0 in (2.2) and having minimum variance among all estimators in this subclass), and (iii) the simple ratio estimator. For $0 \leq K < 1$, nearly optimal values of K and (K, A) are also given (K was incremented by 0.1 and the optimal value of A was determined for each fixed K). Here, a value (K, A) is referred to optimal value if the AK composite estimator with this value has the smallest variance among all AK composite estimators defined by (2.2). Similar definition applies to the term "optimal K". Table 4A (computed using $\hat{\gamma}_j$'s given in Table 3) shows that, for all characteristics except "unemployed" there are 18-21% gains in relative efficiency for the K composite estimates and 26-30% gains in the relative efficiency for the AK composite estimates.

To determine the effect of γ_j's on the relative efficiencies, γ_j's were replaced by zero's in the expression for $V(y'_m)$ and the optimal K, optimal (K, A), and the relative efficiencies were computed. These results are presented in Table 4B. Note that the optimal K's and optimal (K, A)'s in the Tables 4A and 4B are different. Comparison of the corresponding relative efficiencies in these two tables shows that positive γ's have a negative effect on the reduction in variance, i.e., the gains in relative efficiency are reduced. The greatest reduction in relative efficiency is for the characteristic "employed agriculture". This is the characteristic with relatively high values of $\hat{\gamma}_j$'s. Thus taking γ_j's to be zero, when $\gamma_j > 0$, can result in over-estimation of the relative efficiencies and the degree of over-estimation depends on the magnitude of γ_j's.

As mentioned in the introduction, one of the drawbacks of the composite estimators of level is their bias as compared to the simple ratio estimator. Thus comparing the variances of biased estimators can sometimes result in erroneous conclusions about the relative performance of these estimators. It is appropriate to examine the mean square error in the case of biased estimators. The expression for the bias of y'_m (see Appendix I) involves α_i's

(the rotation group biases). The quantity $\hat{\alpha}_i = y_{m,i} - \hat{Y}_m$ is an unbiased estimator of α_i if $\hat{Y}_m$ is an unbiased estimator of Y_m. We assume that the simple ratio estimator $\bar{y}_m$ is an unbiased estimator of Y_m, i.e., $\sum_{i=1}^{6} \alpha_i = 0$. Values of $\hat{\alpha}_i$ ($i = 1,2,\ldots,6$) for various characteristics are given in Table 5. For each of three characteristics "in labour force", "employed" and "employed non-agriculture", we note that: (i) $\hat{\alpha}_1$ is negative while all other $\hat{\alpha}_i$'s are positive; and (ii) $\hat{\alpha}_1$ is large relative to the other $\hat{\alpha}_i$'s.

Table 6 gives the values of optimal K, the optimal (K, A) and results of comparing mean square errors. The optimal K was determined among 10 values of K = 0(0.1)0.9 in the same manner for Tables 4A and 4B. However, the optimal (K, A) was computed in a different way. It was chosen among all possible combinations of K = 0(0.1)0.9 and A = 0(0.1)1.0 rather than determining optimal A for each fixed K = 0(0.1)0.9 (as used for Tables 4A and 4B). Two criteria of optimality are used. One is based on the concept of minimum variance (as is the case for Tables 4A and 4B), and the other is based on the concept of minimum mean square error.

It is shown in Appendix I that

$$E(y_m') = Y_m + [A\alpha_1 + K(\alpha_6 - \alpha_1)]/[5(1-K)].$$

Bias of each estimate in Table 6 is computed by using $\hat{\alpha}_1$ and $\hat{\alpha}_6$ (given in Table 5) instead of α_1 and α_6 in the above formula. Now we discuss the results of Table 6.

For the K composite estimate (based on minimum mean square error optimality) there is only a moderate gain in relative efficiency for the characteristic "employed agriculture" and a nominal gain for the characteristic "unemployed". Also, the bias of the estimates for these two characteristics is small. For the remaining characteristics, the simple ratio estimate is the optimal K composite estimate.

The K composite estimates (considered in Table 4A and based on minimum variance optimality) for the three characteristics "in labour force", "employed" and "employed non-agriculture" have relative efficiencies less than 10%. In these cases, the poor performance can be attributed to the large bias. For each of the remaining two characteristics, K composite estimate is only marginally better than the simple ratio estimate, i.e., the gain in relative efficiency is insignificant. The difference in the corresponding relative efficiency results in Tables 4A and 6 is due to the different relative efficiency definitions used for the two tables. For Table 4A, relative efficiency is defined as the ratio of appropriate variances whereas for Table 6, mean square errors are used instead of the variances.

The AK composite estimate (based on minimum mean square optimality) shows relative efficiency gains in the range 16-22% for all characteristics except "unemployed". Also, the bias of estimate for each characteristic is small.

However, the AK composite estimate based on minimum variance optimality, like the corresponding K composite estimate, has very low relative efficiency for the characteristics "in labour force", "employed", "employed non-agriculture" because of large bias in these cases. The gain in relative efficiency for the characteristic "employed agriculture" is moderate whereas the corresponding gain the characteristic "unemployed" is nominal.

The results in Table 6 show that, among the four composite estimates discussed above, the optimal AK composite estimates (based on minimum mean square error) have relative efficiencies higher for all characteristics than the corresponding relative efficiencies for other composite estimates. We will discuss later the results in the last column of Table 6.

We note, from the expression for $E(y'_m)$ given earlier, the $y'_m - y'_{m-1}$ is an unbiased estimator of $Y_m - Y_{m-1}$, i.e., K or AK composite estimators of change are unbiased. Table 7 gives the optimal K, optimal (K, A), and relative efficiency results for optimal K composite and optimal AK composite estimates of change. The gains in relative efficiency for the characteristics "in labour force", "employed", and "employed non-agriculture" are in the 46-55%

range for K composite and AK composite estimates. For the characteristic "employed agriculture", the optimal AK composite estimate is also optimal K composite and the gain in relative efficiency is about 135%. The gain in relative efficiency for the characteristic "unemployed" is about 6% for both estimates.

It should be pointed out that the optimal value of K or (K, A) is characteristic dependent. Thus the additive property of the estimates is not preserved when different values of K or (K, A) are employed. To preserve additivity, a common value of K = 0.4 and A = 0.4 was selected for estimates of level and change. The following remarks describe the performance of the AK composite estimate with K = 0.4 and A = 0.4. The last column of Table 6 shows that the gains in relative efficiency for AK composite estimates of level are in the 6-10% range for all characteristics except "unemployed". The results of Table 7 show that the gains in relative efficiency for AK composite estimates of change are in the 12-15% range for all characteristics except "unemployed". The gain in relative efficiency for AK composite estimates of level and change is about 2-3% for the characteristic "unemployed".

ACKNOWLEDGEMENT

The authors wish to thank D. Drew, G.B. Gray and M.P. Singh of Statistics Canada for helpful discussions during the progress of this work, and the referee whose suggestions led to many improvements.

TABLE 1

Common and Predecessor Panels Pertaining
To Months m and m-j

Panels in Month m	m-1	m-2	m-3	m-4	m-5	m-6	m-7	m-8	m-9	m-10	m-11
1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))	((4))	((3))	((2))
2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))	((4))	((3))
3	2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))	((4))
4	3	2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))
5	4	3	2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))
6	5	4	3	2	1	(6)	(5)	(4)	(3)	(2)	(1)

The correlation coefficients between common panels of months m and m-j indicated by panels with no parentheses equal ρ_j.

The correlation coefficient between panels of month m and their "single" predecessor of month m-j equals γ_j, the panels indicated by single parentheses.

The correlation coefficient between panels of month m and their double predecessor of month m-j equals $\gamma_j^{(2)}$, the panels indicated by double parentheses. In this report, all $\gamma_j^{(2)}$ are assumed to equal 0.

TABLE 2

Estimated Correlation ρ's (1980-1981 Ontario)

Characteristics \ $\hat{\rho}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	$\hat{\rho}_5$
In Labour Force	.843	.782	.717	.674	.631
Employed	.852	.779	.709	.664	.619
Employed Agriculture	.955	.926	.901	.861	.821
Employed Non-Agriculture	.861	.791	.724	.678	.632
Unemployed	.580	.445	.334	.286	.238

TABLE 3

Estimated Correlation γ's (1980-1981 Ontario)

Characteristics \ $\hat{\gamma}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$	$\hat{\gamma}_7$	$\hat{\gamma}_8$	$\hat{\gamma}_9$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{11}$
In Labour Force	.161	.141	.128	.133	.135	.136	.125	.127	.124	.122	.127
Employed	.164	.136	.142	.142	.146	.149	.148	.150	.153	.141	.148
Employed Agriculture	.477	.483	.474	.486	.480	.474	.459	.429	.394	.323	.252
Employed Non-Agriculture	.184	.150	.147	.157	.162	.167	.166	.169	.174	.156	.166
Unemployed	.141	.074	.076	.063	.057	.051	.045	.060	.077	.136	.074

TABLE 4A & 4B

The Optimal (K, A) and K, and the Relative Efficiencies of K Composite and AK Composite Estimators.

TABLE 4A

$\gamma_i \neq 0$

Characteristics	K composite		AK composite		
	Optimal K	Relative Efficiency	Optimal K	A	Relative Efficiency
In Labour Force	0.7	118.8	0.8	0.48	128.4
Employed	0.7	118.5	0.8	0.49	128.1
Employed Agriculture	0.8	120.6	0.8	0.38	126.9
Employed Non-Agriculture	0.7	119.4	0.8	0.47	129.3
Unemployed	0.3	102.8	0.5	0.38	105.2

TABLE 4B

$\gamma_i = 0$ for all i.

Characteristics	K composite		AK composite		
	Optimal K	Relative Efficiency	Optimal K	A	Relative Efficiency
In Labour Force	0.7	125.5	0.8	0.50	138.4
Employed	0.7	125.3	0.8	0.51	137.9
Employed Agriculture	0.8	167.3	0.9	0.46	187.9
Employed Non-Agriculture	0.7	126.9	0.8	0.49	140.2
Unemployed	0.4	104.4	0.6	0.51	108.4

Relative efficiency is with respect to the simple ratio estimator and is defined as 100 times the ratio V(simple ratio estimator)/V(K or AK composite).

TABLE 5

Estimates (in thousands) of Rotation Group Bias α_i

Characteristics	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_6$
In Labour Force	-135.3	39.8	41.1	31.1	15.4	7.9
Employed	-141.7	35.5	34.9	31.3	25.4	14.8
Employed Agriculture	-4.2	-2.6	2.2	-0.1	4.2	0.5
Employed Non-Agriculture	-137.5	38.0	32.7	31.3	21.2	14.3
Unemployed	6.4	4.3	6.2	-0.1	-9.9	-6.9

α_i is defined as $E(y_{m,i}) - Y_m$ and estimated by $y_{m,i} - \sum_{i=1}^{6} y_{m,i}/6$.

TABLE 6

Comparison of the Variances and the Mean Square Errors of Simple, K Composite, and AK Composite Estimators

IN LABOUR FORCE

	Simple Ratio Estimator	K Composite		AK Composite		Common K, A for all Characteristics
		(Min MSE) K=0	(Min Var) K=0.7	(Min MSE) K=0.7 A=0.7	(Min Var) K=0.8 A=0.5	K=0.4 A=0.4
Monthly Level Estimate 10^3	4480.7	4480.7	4547.5	4484.4	4527.6	4481.7
Variance 10^6	432.0	432.0	363.8	358.1	336.5	391.8
Bias 10^3	0	0	66.8	3.7	46.9	1.0
Mean Square Error 10^6	432.0	432.0	4284.5	371.5	2532.4	392.9
Relative Efficiency		100.0	9.0	116.3	17.1	110.0

EMPLOYED

		K=0	K=0.7	K=0.8 A=0.9	K=0.8 A=0.5	K=0.4 A=0.4
Monthly Level Estimate 10^3	4186.0	4186.0	4259.0	4183.6	4240.3	4188.0
Variance 10^6	473.3	473.3	399.6	397.7	369.5	428.9
Bias 10^3	0	0	73.0	-2.4	54.3	2.0
Mean Square Error 10^6	473.3	473.3	5732.2	403.2	3320.9	432.8
Relative Efficiency		100.0	8.3	117.4	14.3	109.4

EMPLOYED AGRICULTURE (TABLE 6 continued)

	Simple Ratio Estimator	K Composite		AK Composite		Common K, A for all Characteristics
		(Min MSE) K=0.6	(Min Var) K=0.8	(Min MSE) K=0.8 A=0.6	(Min Var) K=0.8 A=0.4	K=0.4 A=0.4
Monthly Level Estimate 10^3	142.0	143.4	145.7	143.2	144.1	142.1
Variance 10^6	85.7	75.6	71.1	68.7	67.6	80.8
Bias 10^3	0	1.4	3.7	1.2	2.1	0.1
Mean Square Error 10^6	85.7	77.6	85.1	70.2	71.8	80.8
Relative Efficiency		110.5	110.7	122.2	119.4	106.1

EMPLOYED NON-AGRICULTURE

		K=0	K=0.7	K=0.8 A=0.9	K=0.8 A=0.5	K=0.4 A=0.4
Monthly Level Estimate 10^3	4043.9	4043.9	4114.7	4041.6	4096.6	4045.8
Variance 10^6	498.9	498.9	417.8	418.0	385.9	452.8
Bias 10^3	0	0	70.8	-2.3	52.7	1.9
Mean Square Error 10^6	498.9	498.9	5436.1	423.3	3161.7	456.4
Relative Efficiency		100.0	9.2	117.9	15.8	109.3

UNEMPLOYED (TABLE 6 continued)

	Simple Ratio Estimator	K Composite		AK Composite		Common K, A for all Characteristics	
		(Min MSE) K=0.2	(Min Var) K=0.3	(Min MSE). K=0.4 A=0.4	(Min Var) K=0.5 A=0.4	K=0.4	A=0.4
Monthly Level Estimate 10^3	294.8	294.1	293.7	293.9	293.2	293.9	
Variance 10^6	117.5	114.9	414.3	112.5	111.7	112.5	
Bias 10^3	0	-0.7	-1.1	-0.9	-1.6	-0.9	
Mean Square Error 10^6	117.5	115.4	115.7	113.3	114.4	113.3	
Relative Efficiency		101.9	101.6	103.7	102.7	103.7	

Relative efficiency is relative to the simple ratio estimator and is defined by 100 times MSE(simple ratio estimator)/MSE(K or AK Composite estimator).

TABLE 7

Relative Efficiency of Composite Estimators
for Month-to-Month Change

Labour Force Characteristics	K composite		AK Composite			Common K, A K=0.4 A=0.4
	Optimal K	Relative Efficiency	Optimal K	A	Relative Efficiency	Relative Efficiency
In Labour Force	0.9	146.6	0.9	0.1	147.9	113.3
Employed	0.9	151.0	0.9	0.1	152.3	114.1
Employed Agriculture	0.9	234.7	0.9	0.0	234.7	112.3
Employed Non-Agriculture	0.9	154.0	0.9	0.1	155.2	114.1
Unemployed	0.4	106.0	0.6	0.2	106.4	102.9

Relative efficiency is with respect to the simple ratio estimator and is defined as 100 times the ratio of appropriate variances.

APPENDIX I

Derivation of Bias of the Composite Estimator:

As defined in (2.2), the AK composite estimator of Y_m is given by:

$$y'_m = (1 - K + A)y_{m,1}/6 + (1 - K - A/5)\sum_{j=2}^{6} y_{m,j}/6 + K(y'_{m-1} + d_{m,m-1}). \quad (A1.1)$$

It may be noted that the simple ratio estimator now employed in LFS is the average of the six panel estimators and is given by:

$$\bar{y}_m = \sum_{i=1}^{6} y_{m,i}/6. \quad (A1.2)$$

From (2.3), the bias of the i-th panel estimator equals α_i so that:

$E(y_{m,i}) = Y_m + \alpha_i$, recalling that the bias is independent of m. Hence,

$$E(\bar{y}_m) = Y_m + \sum_{i=1}^{6} \alpha_i/6 = Y_m + \bar{\alpha} \text{ (say)}.$$

In later developments we assume that $\bar{\alpha} = 0$.

The composite estimator may be rewritten as:

$$y'_m = y_m + K(y'_{m-1} + d_{m,m-1}), \tag{A1.3}$$

where

$$y_m = (1 - K + A)y_{m,1}/6 + (1 - K - A/5) \sum_{j=2}^{6} y_{m,j}/6$$

$$= (1 - K)\bar{y}_m + A(y_{m,1} - \bar{y}_m)/5.$$

Therefore

$$E(y_m) = (1 - K)(Y_m + \bar{\alpha}) + (A/5)(\alpha_1 - \bar{\alpha}). \tag{A1.4}$$

When $\bar{\alpha} = 0$, it simplifies to

$$E(y_m) = (1 - K)Y_m + (A/5)\alpha_1.$$

Using the definition of $d_{m,m-1}$ given in (2.1), we have

$$E(d_{m,m-1}) = E[\sum_{j=2}^{6} (y_{m,j} - y_{m-1,j-1})/5]$$

$$= (Y_m - Y_{m-1}) + (\alpha_6 - \alpha_1)/5. \tag{A1.5}$$

Now y'_m may be expanded by applying (A1.3) recursively and it is found that, up to n months back:

$$y'_m = y_m + Ky_{m-1} + K^2y_{m-2} + \ldots + K^{n-1}y_{m-n+1} + K^n y'_{m-n}$$

$$+ Kd_{m,m-1} + K^2 d_{m-1,m-2} + \ldots + K^n d_{m-n+1,m-n}. \tag{A1.6}$$

The expected value of y'_m may be readily obtained from (A1.4) and (A1.6) as below:

$$E(y'_m) = (1 - K)\,[Y_m + KY_{m-1} + K^2Y_{m-2} + \ldots + K^{n-1}Y_{m-n+1}] + K^nE(y'_{m-n})$$

$$+ [(1 - K)\,\bar{\alpha} + (A/5)(\alpha_1 - \bar{\alpha})](1 - K^n)/(1 - K)$$

$$+ K(Y_m - Y_{m-1}) + K^2(Y_{m-1} - Y_{m-2}) + \ldots + K^n(Y_{m-n+1} - Y_{m-n})$$

$$+ [(\alpha_6 - \alpha_1)/5]K(1 - K^n)/(1 - K)$$

$$= Y_m + K^n[E(y'_{m-n}) - Y_{m-n})]$$

$$+ [(1 - K)\bar{\alpha} + (A/5)(\alpha_1 - \bar{\alpha}) + K(\alpha_6 - \alpha_1)/5](1 - K^n)/(1 - K)$$

$$= Y_m + K^n[E(y'_{m-n}) - Y_{m-n}]$$

$$+ [(1 - K - A/5)\bar{\alpha} + (A/5)\alpha_1 + K(\alpha_6 - \alpha_1)/5](1 - K^n)/(1 - K) \qquad (A1.7)$$

which simplifies for sufficiently large n and for the case $\bar{\alpha} = 0$ to

$$E(y'_m) = Y_m + [A\alpha_1 + K(\alpha_6 - \alpha_1)]/[5(1 - K)]. \qquad (A1.8)$$

Since the bias of y'_m under the model assumed in this paper is independent of m, the difference between composite estimates r months apart is unbiased, i.e.,

$$E(y'_m - y'_{m-r}) = Y_m - Y_{m-r} \text{ for all r.} \qquad (A1.9)$$

APPENDIX II

Derivation of the Variance of the Composite Estimator

We assume that the composite estimators (see (2.2)) have become sufficiently stable over time and hence we shall assume that $V(y'_{m-1}) = V(y'_m)$. Since all correlations 12 or more months apart are assumed to be zero, we shall assume that the LFS composite estimators have become stable after 12 months.

Taking the variance of both sides of (A1.3) and applying the above assumption one may solve for $V(y'_m)$ to find that:

$$V(y'_m) = [V(y_m) + K^2V(d_{m,m-1}) + 2KCov(y_m, d_{m,m-1}) + 2KCov(y_m, y'_{m-1}) + 2K^2Cov(d_{m,m-1}, y'_{m-1})]/(1 - K^2). \quad (A2.1)$$

To eliminate y'_{m-1} on the right side of (A2.1), we apply (A1.6), replacing m by (m - 1) and n by 12 to obtain:

$$y'_{m-1} = \sum_{g=1}^{12} (K^{g-1}y_{m-g} + K^g d_{m-g,m-g-1}) + K^{12}y'_{m-13}. \quad (A2.2)$$

Substituting (A2.2) in (A2.1) and dropping zero terms, we have

$$V(y'_m) = \{V(y_m) + K^2V(d_{m,m-1}) + 2KCov(y_m, d_{m,m-1}) + 2\sum_{q=1}^{12} K^g[Cov(y_m, y_{m-g}) + KCov(d_{m,m-1}, y_{m-g}) + KCov(y_m, d_{m-q,m-g-1}) + K^2Cov(d_{m,m-1}, d_{m-g,m-g-1})]\}/(1 - K^2). \quad (A2.3)$$

We give the expressions for the variances and covariances on the right side of (A2.3), which may be readily derived by considering (2.1) and (A1.3).

$$V(y_m) = [(1 - K)^2/6 + A^2/30]\sigma^2, \quad (A2.4)$$

which simplifies to $\sigma^2/6$ when A = K = 0; i.e.,

$$V(\bar{y}_m) = \sigma^2/6, \quad (A2.4a)$$

the variance of the current LFS estimator.

$$V(d_{m,m-1}) = 2\sigma^2(1 - \rho_1)/5. \qquad (A2.5)$$

$$Cov(y_m, d_{m,m-1}) = (1 - K)(1 - \rho_1)\sigma^2/6 - A(1 - \rho_1)c^2/30. \qquad (A2.6)$$

To derive the remaining covariances in (A2.3), which involve 'g', an indicator function I(a, b) shall be defined by:

$$I(a, b) = 1 \text{ if } a \leq b$$
$$= 0 \text{ otherwise.}$$

By considering the definitions of y_m in (A1.3), $d_{m,m-1}$ in (2.1) and the corresponding expressions for month (m - q), one would find that the following covariances would be required to derive the remaining covariance of (A2.3).

$$Cov(\bar{y}_m, \bar{y}_{m-q}) = (\sigma^2/36)[(6 - q)\rho_q I(q, 5) + (6 - |\gamma - 6|)\gamma_q],$$

$$Cov(\bar{y}_m, y_{m-q,1}) = (\sigma^2/6)[\rho_q I(q, 5) + \gamma_q I(6, q)],$$

$$Cov(y_{m,1}, \bar{y}_{m-q}) = (\sigma^2/6)\gamma_q I(q, 6),$$

$$Cov(y_{m,1}, y_{m-q,1}) = \sigma^2 I(q, 6)I(6, q)\gamma_6 \ (= 0 \text{ except when } q = 6),$$

$$Cov(y_{m-1,6}, \bar{y}_{m-q}) = (\sigma^2/6)[\rho_{q-1} I(q, 6) + \gamma_{q-1} I(7, q)],$$

$$Cov(y_{m-1,6}, y_{m-q,1}) = \sigma^2[\rho_{q-1} I(q, 6)I(6, q)],$$

$$Cov(\bar{y}_m, y_{m-q-1,6}) = (\sigma^2/6)\gamma_{q+1} I(q, 5),$$

$$Cov(y_{m,1}, y_{m-q-1,6}) = \sigma^2\gamma_1 I(q, 0)I(0, q) \ (= 0 \text{ for } q \geq 1). \qquad (A2.7)$$

The four covariances of (A2.3) that involve q may be readily defined and are found to be as follows:

$$Cov(y_m, y_{m-q}) = \sigma^2\rho_q I(q, 5)[(1 - K)^2(6 - q)/36$$

$$+ A(1 - K)(g - 3)/90 - gA^2/900]$$

$$+ \sigma^2\gamma_g[(1 - K)^2(6 - |g - 6|)/36$$

$$+ A(1 - K)(|g - 6| - 3)/90 - |g - 6| A^2/900]$$

$$+ \sigma^2\gamma_g I(g, 6)I(6, g)A(1 - K + A)/30, \quad (A2.8)$$

$$\mathrm{Cov}(d_{m,m-1}, y_{m-g}) = \sigma^2(\rho_g - \rho_{g-1})I(g, 5)[(1 - K)(6 - g)/30 + gA/150]$$

$$+ \sigma^2(\gamma_g - \gamma_{g-1})[(1 - K)(6 - |g - 6|)/30 + |g - 6|A/150$$

$$- (1 - K + A)I(g, 6)/30], \quad (A2.9)$$

$$\mathrm{Cov}(y_m, d_{m-g,m-g-1}) = \sigma^2(\rho_g - \rho_{g+1})I(g, 5)(1 - K - A/5)(5 - g)/30$$

$$+ \sigma^2(\gamma_g - \gamma_{g+1})[(1 - K - A/5)(6 - I(6, g)$$

$$- |g - 6|)/30 + AI(g, 5)/25], \quad (A2.10)$$

$$\mathrm{Cov}(d_{m,m-1}, d_{m-g,m-g-1}) = \sigma^2[(5 - g)(2\rho_g - \rho_{g-1} - \rho_{g+1})I(g, 5)$$

$$+ (5 - |g - 6|)(2\gamma_g - \gamma_{g-1} - \gamma_{g+1})]/25. \quad (A2.11)$$

Hence, $V(y'_m)$ can be expressed as $aA^2 + bA + c = f(A)$ where a, b and c are functions of K, ρ's and γ's. It can be shown that $a \geq 0$. The values of A that minimize the variance of AK estimator was determined for K = 0(0.1)0.9. Among these (A, K)'s, the optimal value of (A, K) was selected and is presented in Table 4A.

APPENDIX III

Derivation of the variance of $Y'_m - Y'_{m-1}$

From (A1.3)

$$y'_m = y_m + K(y'_{m-1} + d_{m,m-1}), \text{ or}$$

$$y'_m - Ky'_{m-1} = y_m + Kd_{m,m-1},$$

whence

$$(1 + K^2)V(y'_m) - 2KCov(y'_m, y'_{m-1}) = V(y_m) + 2KCov(y_m, d_{m,m-1}) + K^2V(d_{m,m-1}).$$

When $K \neq 0$, $Cov(y'_m, y'_{m-1})$ may be obtained from the above and upon substitution of (A2.4), (A2.6) and (A2.5), and from the fact that $V(y'_m - y'_{m-1}) = 2V(y'_m) - 2Cov(y'_m, y'_{m-1})$, one may find that for $K \neq 0$:

$$V(y'_m - y'_{m-1}) = \sigma^2[A^2/30 - (1 - \rho_1)KA/15 + (1 - K)^2/6 + (1 - \rho_1)K(K + 5)/15]/K - (1 - K)^2V(y'_m)/K. \quad (A3.1)$$

When $K = 0$,

$$y'_m = (1 - A/5)\bar{y}_m + Ay_{m,1}/5,$$

$$Cov(y'_m, y'_{m-1}) = Cov[(1 - A/5)\bar{y}_m + Ay_{m,1}/5, (1 - A/5)\bar{y}_{m-1} + Ay_{m-1,1}/5]. \quad (A3.2)$$

Thus for $K = 0$, we have:

$$V(y'_m - y'_{m-1}) = \sigma^2[(1/15 + \rho_1/450 + \gamma_1/90)A^2 + 2(\rho_1 - \gamma_1)A/45 + 1/3 - (5\rho_1 + \gamma_1)/18]. \quad (A3.3)$$

REFERENCES

[1] Bailar, B.A. (1975), "The Effects of Rotation Group Bias on Estimates from Panel Surveys", Journal of the American Statistical Association,

70, 23-29.

[2] Ghangurde, P.D. (1982), "Rotation Group Bias in the LFS Estimates". Survey Methodology, 8, 86-101.

[3] Gurney, M., and Daly, J.F. (1965), "A Multivariate Approach to Estimation in Periodic Sample Surveys", Proceedings of the Social Statistics Section, American Statistical Association, 242-257.

[4] Huang, E., and Ernst, L. (1981), "Comparison of an Alternate Estimator to the Current Composite Estimator in CPS", Proceedings of the Section on Survey Research Methods, American Statistical Association, 303-308.

[5] Lee, H. (1983), "Estimation of Panel Correlations in the Canadian Labour Force Survey", Technical Memorandum (in Preparation), Census and Household Survey Methods Division, Statistics Canada.

[6] Tessier, R. and Tremblay, V. (1976), "Findings on Rotation Group Biases". Technical Memorandum, Statistics Canada.

[7] Statistics Canada (1976), Methodology for the Canadian Labour Force Survey, Catalogue 71-526, Occasional.

Statistics Canada Statistique Canada

Methodology Branch

Census & Household Survey
Methods Division

Direction de la méthodologie

Division des méthodes de recensement
et d'enquêtes-ménages

Canada

S-176F
C.2.

EVALUATION DE L'APPLICATION D'ESTIMATEURS COMPOSITES À L'ENQUETE SUR LA POPULATION ACTIVE DU CANADA[1]

S. Kumar et H. Lee[2]

Division des méthodes de recensement et d'enquêtes-ménages

Statistique Canada

Numéro: DMREM 85-067F

ÉVALUATION DE L'APPLICATION D'ESTIMATEURS COMPOSITES À L'ENQUÊTE SUR LA POPULATION ACTIVE DU CANADA[1]

S. Kumar et H. Lee[2]

Cette étude porte sur l'applicabilité des techniques d'estimation composite à l'enquête sur la population active du Canada. On examine l'efficacité d'une classe d'estimateurs composites AK proposés en premier par Gurney et Daly dans l'estimation de plusieurs caractéristiques de la population active. L'estimateur composite ordinaire présente un biais considérable, mais l'estimateur composite AK permet de réduire ce biais. On compare des estimateurs composites à variance minimum et à erreur quadratique moyenne minimum.

1. INTRODUCTION

L'enquête sur la population active du Canada (EPA) est menée tous les mois par Statistique Canada pour produire des estimations de diverses caractéristiques de la population active. Le plan de sondage de l'EPA comprend un plan de renouvellement qui permet de remplacer un sixième des ménages de l'échantillon à chaque mois (voir [7]). L'échantillon est ainsi composé de six panels ou groupes de renouvellement et chaque panel est dans l'échantillon pendant six mois consécutifs.

Comme Bailar [1] l'a déjà souligné, un des principaux inconvénients de l'estimateur composite actuellement utilisé pour la U.S. Current Population Survey (CPS) en comparaison de l'estimateur simple par quotient est son biais. Ce biais découle des différences entre les groupes de renouvellement; c'est-à-dire que les estimations calculées à partir de données relatives à différents

---

[1] Exposé présenté aux "Joint Statistical Meetings of the American Statistical Association, the Biometric Society, the Institute of Mathematical Statistics et the Statistical Society of Canada", Toronto, août 1983.

[2] S. Kumar et H. Lee, Division des méthodes de recensement et d'enquêtes-ménages, Statistique Canada.

groupes de renouvellement au cours d'une même période peuvent ne pas avoir la même espérance mathématique. Ce phénomène, qu'on appelle souvent le biais de renouvellement, a déjà été étudié pour l'EPA (voir [2] et [6]). Récemment, Huang et Ernst [4] ont décrit les résultats de l'application à la CPS de l'estimateur composite AK, qui a été proposé par Gurney et Daly [3] (A et K sont des constantes dans l'équation qui définit cet estimateur composite). Leurs résultats indiquent que l'estimateur AK est supérieur aux estimateurs composites actuellement utilisés dans la CPS du point du vue de la variance et du biais.

La présente étude a pour objet d'évaluer l'applicabilité des techniques d'estimation composite à l'EPA. Les résultats de l'utilisation de différents estimateurs composites pour le calcul de totaux et de changements sont examinés dans le cas de cinq caractéristiques, soit la taille de la population active, le nombre de personnes occupées, le nombre de personnes occupées dans l'agriculture, le nombre de personnes occupées dans les secteurs non agricoles, et le nombre de chômeurs. Les estimations composites de ces caractéristiques sont comparées avec l'estimation simple par quotient qui est utilisée à l'heure actuelle dans l'EPA. Cette étude est basée sur les données de 1980 à 1981 concernant la province de l'Ontario.

2. DÉFINITIONS ET NOTATION

Nous voulons estimer Y_m le nombre de personnes dans la population qui ont une certaine caractéristique au mois m. Définissons les variables suivantes :

$y_{m,i}$ = un estimateur simple par quotient de Y_m calculé pour le i^e panel ($i = 1,2,\ldots,6$). Ici le i^e panel correspond au sous-échantillon (groupe de renouvellement) qui fait partie de l'échantillon pour la i^e fois. On appelera ici cette variable l'estimateur relatif au i^e panel.

$d_{m,m-1}$ = l'estimateur du changement ($Y_m - Y_{m-1}$) survenu du mois (m - 1) au mois m dans l'ensemble des cinq panels qui faisaient partie de

l'échantillon à la fois au mois m et au mois (m - 1)

$$= \sum_{j=2}^{6} (y_{m,j} - y_{m-1.j-1})/5 \qquad (2.1)$$

y'_m = l'estimateur composite AK de Y_m défini par l'équation

$$y'_m = (1 - K + A)y_{m,1}/6 + (1 - K - \frac{A}{5}) \sum_{j=2}^{6} y_{m,j}/6$$

$$+ K(y'_{m-1} + d_{m,m-1}) \qquad (2.2)$$

où K et A sont des constantes et $0 \leq K < 1$.

L'équation (2.1) définit une classe d'estimateurs qui portent le nom d'estimateurs composites AK. Les estimateurs qu'on obtient lorsque A = 0 dans la formule (2.2) constituent des estimateurs composites K. L'estimateur simple par quotient, qu'on représente ici par $\bar{y}_m$, la moyenne des estimateurs relatifs aux six panels, est exprimé par la forme de l'équation (2.2) dans laquelle A = 0 et K = 0. Nous comparons plus bas les résultats de l'utilisation de l'estimateur composite AK optimal (selon un critère de variance minimum ou d'erreur quadratique moyenne minimum), de l'estimateur composite K et de l'estimateur simple par quotient.

Nous supposons que le biais de renouvellement $E(y_{m,i}) - Y_m$ est indépendant de m et qu'il varie en fonction de i. Ce biais est représenté par le symbole α_i.

Algébriquement,

$$\alpha_i = E(y_{m,i}) - Y_m. \qquad (2.3)$$

Le calcul du biais de l'estimateur composite est présenté à l'annexe I.

3. HYPOTHÈSES

Le plan de renouvellement de l'échantillon de l'EPA peut être représenté schématiquement à l'aide du tableau 1, où au mois m le panel i = (1,2....,6), (qui est à son i^e mois d'inclusion dans l'échantillon) est le même que le panel (i - j) au mois (m - j), où $1 \leq (i - j) \leq 5$. Le panel i au mois m a été précédé immédiatement par le panel (6 + i - j) au mois (m - j), où $1 \leq (6 + i - j) \leq 6$. De même, le panel i au mois m a été précédé en deuxième lieu par le panel (12 + i - j) au mois m - j, où $1 \leq (12 + i - j) \leq 6$. En général, le r^e prédécesseur du panel i au mois m correspond au panel (i - j + 6r) au mois m - j. Notons que le 0^e prédécesseur du panel i représente le même panel de mois précédents.

L'expression de la variance de y'_m, c'est-à-dire $V(y'_m)$, est une fonction des variances et des covariances des divers estimateurs relatifs aux panels (voir l'annexe II). On suppose que la structure de ces variances et covariances a les propriétés suivantes, qui concordent avec le plan de renouvellement de l'EPA qui est résumé au tableau 1.

(i) $V(y_{m,i}) = \sigma^2$ pour tout m, i = 1,2,...,6

(ii) $Cov(y_{m,i}, y_{m-j,i-j+6r}) = \gamma_j^{(r)}\sigma^2$, où i = 1,2,...,6, j > 0 et $r \geq 0$, de sorte que $6 \geq i - j + 6r \geq 1$. Ici r représente le nombre de prédécesseurs du panel courant.

Pour r = 0, c'est-à-dire que $6 > i - j \geq 1$, soit $\gamma_j^{(r)}\sigma^2 = \rho_j$ (cas des panels inclus dans l'échantillon au mois m et au mois m-j).

Pour r = 1, c'est-à-dire que $6 \geq i - j + 6 \geq 1$, soit $\gamma_j^{(r)} = \gamma_j$ (cas du panel courant et de son prédécesseur immédiat j mois plus tôt).

Pour $r \geq 2$, c'est-à-dire que $6 \geq i - j + 6r \geq 1$, soit $\gamma_j^{(r)} = 0$ (cas du panel courant et de son r^e prédécesseur j mois plus tôt).

(iii) Le développement mathématique de la variance de l'estimateur compo-

site y_m doit tenir compte des coefficients de corrélation ρ_j et γ_j, qui sont tous les deux supposés stationnaires: autrement dit, ces coefficients sont des fonctions de j et non de m. Il est raisonnable de supposer que les ρ_j et les γ_j sont tous positifs puisque les ρ_j correspondent à des caractéristiques de ménages qui sont dans l'ensemble les mêmes, tandis que les γ_j reflètent les caractéristiques de ménages inclus dans l'échantillon au mois courant et celles de leurs proches voisins (dans bien des cas leurs voisins de quartier) j mois plus tôt (sans mentionner l'effet du renouvellement des grappes).

(iv) L'expression de $V(y'_m)$ comprend des termes de covariance qui ne font pas partie des hypothèses (ii) et (iii), comme, par exemple,

$Cov(y_{m,i}, y_{m,j})$ pour $i \neq j$, $Cov(y_{m,i}, y_{m-1,j})$ pour $i = 1$, $j \neq 6$ et $i \neq 1$, $j \neq i - 1$ et $Cov(y_{m,i}, y_{m-q,i})$ pour $q \geq 12$. Ces covariances et toutes les autres qui ne sont pas définies plus haut, y compris celles qui dépendent de $\gamma_j^{(2)}$ et qui sont incluses dans l'expression de $V(y'_m)$, sont supposées nulles.

À partir de ces hypothèses, on a formulé une expression pour la variance de l'estimateur composite AK en fonction des paramètres décrits plus haut. Les raisonnements mathématiques pour l'élaboration des expressions qui déterminent le biais et la variance de y'_m et la variance de $y'_m - y'_{m-1}$ figurent dans les annexes.

4. RÉSULTATS ET ANALYSE

Les variables σ^2 dans l'expression de $V(y'_m)$, σ^2, ρ_j et γ_j, ont été remplacés par les estimations appropriées (pour une description des méthodes d'estimation des ρ et des γ, voir [5]). Notons que, dans l'EPA, les ρ_j n'existent pas pour $j \geq 6$ parce que aucun panel ne remonte si loin. Il n'y a pas non plus de γ pour $j \geq 12$ parce qu'on remonte alors aux prédécesseurs de deuxième rang ou plus du panel courant et on peut supposer que ce genre de corrélation

est nul dans les raisonnements mathématiques. Les estimations de ρ_j, à savoir $\hat{\rho}_j$, sont présentées au tableau 2. On a estimé ρ_5 en extrapolant à partir des autres ρ_j parce qu'il n'était pas possible de l'estimer directement pour l'échantillon. Notons que $\hat{\rho}_j$ ($j = 1,2,\ldots,5$) est une fonction décroissante de j pour toutes les cinq caractéristiques d'intérêt. Ce résultat concorde avec nos intuitions au sujet de l'allure des ρ_j. En outre, les ρ_j ont une valeur élevée pour toutes les caractéristiques sauf le nombre de chômeurs.

Les estimations $\hat{\gamma}_j$, de γ_j, figurent au tableau 3. Les valeurs de $\hat{\gamma}_5$ et $\hat{\gamma}_{11}$ sont estimées respectivement par interpolation et extrapolation à partir d'autres valeurs de $\hat{\gamma}_j$. Intuitivement, on s'attend à voir les γ_j diminuer en fonction de j pour chaque caractéristique Toutefois, on peut constater que ce n'est pas le cas des $\hat{\gamma}_j$. Bien que les $\hat{\gamma}_j$ ne semblent pas être des fonctions monotones décroissantes, on peut signaler que, quand l'écart $(\hat{\gamma}_{j+1} - \hat{\gamma}_j)$ est positif, il est toutefois très petit. Le fait que cet écart puisse être positif est peut-être dû à la variabilité d'échantillonnage plutôt qu'à l'existence d'un écart réellement positif de $(\gamma_{j+1} - \gamma_j)$.

Dans ce qui suit, le terme "efficacité relative" de l'estimateur composite AK (ou K) se rapporte à l'efficacité de cet estimateur par rapport à l'estimateur simple.

Les tableaux 4A et 4B présentent les résultats d'une comparaison des estimations de la variance de trois estimateurs, soit (i) l'estimateur composite AK optimal, c'est-à-dire l'estimateur qui a une variance minimum dans la classe d'estimateurs définis par l'équation (2.2), (ii) l'estimateur composite K optimal (qui correspond à la formule (2.2) avec A = 0 et qui a le minimum de variance dans cette sous-catégorie d'estimateurs) et (iii) l'estimateur simple par quotient. Pour $0 \leq K < 1$, ces deux tableaux indiquent également les valeurs presque optimales de K et de (K, A) (on a balayé cet intervalle en augmentant de 0.1 la valeur de K et la valeur optimale de A a été déterminée en fonction de K). Ici, un couple de valeurs (K, A) est considéré comme étant optimal si l'estimateur composite AK qui le contient a la variance la plus faible de tous les estimateurs composites AK définis par (2.2). Une défini-

tion semblable s'applique dans le cas de la valeur optimale de K. Les chiffres du tableau 4A sur l'efficacité relative (qui ont été calculés à partir des $\hat{\gamma}_j$ du tableau 3) montrent que, pour toutes les caractéristiques sauf le nombre de chômeurs, l'estimateur composite K offre un gain d'efficacité de 18 à 21 % par rapport à l'estimateur simple, tandis que l'estimateur composite AK permet d'accroître l'efficacité de 26 à 30 %.

Pour préciser l'effet des γ_j sur l'efficacité relative, on a annulé les γ_j dans l'expression de $V(y'_m)$ et calculé les valeurs optimales de K et (K, A), de même que l'efficacité relative des estimateurs composites. Les résultats figurent au tableau 4B. Notons que les valeurs optimales de K et (K, A) présentées aux tableaux 4A et 4B sont différentes. Une comparaison des indices correspondants d'efficacité relative dans ces deux tableaux révèle que les γ positifs ont un effet négatif sur la réduction de la variance; autrement dit, l'efficacité relative diminue. La plus grande baisse d'efficacité relative a lieu pour la caractéristique "nombre de personnes occupées dans l'agriculture". C'est cette caractéristique qui a des valeurs relativement élevées de $\hat{\gamma}_j$. Ainsi, si on suppose que les γ_j sont nuls alors qu'en réalité $\gamma_j > 0$, on risque de surestimer l'efficacité relative, et le degré de surestimation dépend de la grandeur des γ_j.

Tel qu'il a été mentionné dans l'introduction, un des inconvénients des estimateurs composites par rapport à l'estimateur simple par quotient est leur biais. Une comparaison des variances d'estimateurs biaisés peut parfois mener à des conclusions erronées au sujet de la qualité relative des estimateurs. Lorsque des estimateurs sont biaisés, il convient d'examiner l'erreur quadratique moyenne. L'expression du biais de y' (voir l'annexe I) comprend les valeurs de α_i (le biais de renouvellement). La variable $\hat{\alpha}_i = y_{m,i} - \hat{Y}_m$ est un estimateur sans biais de α_i si $\hat{Y}_m$ est un estimateur sans biais de Y_m. Nous supposons que l'estimateur simple par quotient $\bar{y}_m$ est un estimateur sans biais de Y_m, c'est-à-dire $\sum_{i=1}^{6} \alpha_i = 0$. Les valeurs de $\hat{\alpha}_i$ ($i = 1,2,\ldots,6$) pour les diverses caractéristiques sont présentées au tableau 5. On constate, pour chacune des trois caractéristiques "taille de la population active", "nombre de personnes occupées" et "nombre de personnes occupées dans les secteurs non

agricoles" (i) que $\hat{\alpha}_i$ est négatif alors que tous les autres $\hat{\alpha}_i$ sont positifs et (ii) que $|\hat{\alpha}_i|$ est grand en comparaison des autres $\hat{\alpha}_i$.

Le tableau 6 montre les valeurs optimales de K et du couple (K, A) et permet de comparer les erreurs quadratiques moyennes des estimations. La valeur optimale de K a été déterminée parmi les dix valeurs dans l'intervalle K = 0(0.1)0.9, comme aux tableaux 4A et 4B. Toutefois, la valeur optimale de (K, A) a été calculée d'une manière différente. Elle a été choisie parmi toutes les combinaisons possibles dans les intervalles K = 0(0.1)0.9 et A = 0(0.1)1.0 au lieu d'être déterminée en fonction de chaque valeur de K dans l'intervalle K = 0(0.1)0.9 (comme dans le cas des tableaux 4A et 4B). Deux critères d'optimalité ont été utilisés. Le premier est fondé sur la recherche de l'estimateur à variance minimum (comme aux tableaux 4A et 4B) et le deuxième sur le calcul de l'estimateur à erreur quadratique moyenne minimum.

On montre à l'annexe I que

$$E(y'_m) = Y_m + [A\alpha_1 + K(\alpha_6 - \alpha_1)]/[5(1 - K)].$$

Le biais de chaque estimation au tableau 6 est calculé en fonction des valeurs de $\hat{\alpha}_1$ et $\hat{\alpha}_6$ (qu'on trouve au tableau 5) au lieu de α_1 et α_6, qui figurent dans la formule écrite plus haut. Nous abordons maintenant les résultats présentés au tableau 6.

L'estimateur composite K basé sur le critère d'optimalité à erreur quadratique moyenne minimum n'affiche qu'un léger gain d'efficacité par rapport à l'estimateur simple pour la caractéristique "nombre de personnes occupées dans l'agriculture" et un gain infime dans le cas du nombre de chômeurs. En outre, le biais des estimations relatives à ces deux caractéristiques est faible. Quant aux autres estimations, l'estimateur optimal produit le même résultat que l'estimateur K optimal.

L'efficacité relative des estimateurs composites K basés sur le critère d'optimalité à variance minimale (qui sont examinés au tableau 4A) est de moins de

10 % pour les trois caractéristiques "taille de la population active". "nombre de personnes occupées" et "nombre de personnes occupées ailleurs que dans l'agriculture". On peut attribuer ces piètres résultats à l'ampleur du biais. Quant aux deux autres caractéristiques, l'estimateur composite K ne présente qu'un faible avantage en comparaison de l'estimateur simple par quotient; en d'autres termes, le gain d'efficacité offert par l'estimateur K est négligeable. Les différences entre les indices d'efficacité relative des tableaux 4A et 6 sont dues au fait que l'efficacité relative n'est pas définie de la même manière dans ces deux tableaux. Au tableau 4A, l'efficacité relative est définie comme le rapport des variances appropriées, tandis qu'au tableau 6 on utilise les erreurs quadratiques moyennes au lieu des variances.

L'estimateur composite AK basé sur le critère d'optimalité à erreur quadratique moyenne minimum affiche un gain d'efficacité de 16 à 22 % par rapport à l'estimateur simple, pour toutes les caractéristiques sauf le nombre de chômeurs. En outre, le biais de l'estimation relative à chaque caractéristique est faible.

Par contre, l'estimateur composite AK basé sur le critère d'optimalité à variance minimum a, comme l'estimateur composite K, une efficacité relative très basse pour les caractéristiques "taille de la population active", "nombre de personnes occupées" et "nombre de personnes occupées dans les secteurs non agricoles" parce que le biais est assez grand dans ces cas. L'efficacité relative est assez bonne pour la caractéristique "nombre de personnes occupées dans l'agriculture", mais elle dépasse 100 de très peu dans le cas du nombre de chômeurs.

Les résultats montrent que, parmi les quatre estimateurs composites qui figurent au tableau 6, c'est l'estimateur composite AK optimal à erreur quadratique moyenne minimum qui a l'efficacité relative la plus élevée pour toutes les caractéristiques. Il sera question de la dernière colonne du tableau 6 un peu plus loin.

On peut noter, à partir de l'expression de $E(y'_m)$ qui est donnée plus haut, que $y'_m - y'_{m-1}$ est un estimateur sans biais de $Y_m - Y_{m-1}$, ce qui revient à dire que

les estimateurs composites K ou AK permettent de mesurer le changement sans biais. Le tableau 7 présente les valeurs optimales de K et (K, A) et montre l'indice d'efficacité relative des estimateurs composites K et AK relatifs aux changements. Les estimations des caractéristiques "taille de la population active", "nombre de personnes occupées", et "nombre de personnes occupées dans les secteurs non agricoles" affichent des gains d'efficacité de 46 à 55 % dans le cas des estimateurs composites K et AK. Quant au nombre de personnes occupées dans l'agriculture, les estimateurs composites AK et K ont le même résultat et la marge d'efficacité relative est d'environ 135 %. Ces deux estimateurs offrent un gain d'efficacité d'à peu près 6 % pour la caractéristique "nombre de chômeurs".

Il est important de souligner que la valeur optimale de K ou (K, A) varie en fonction de la caractéristique estimée. Or, la propriété additive des estimateurs n'est pas conservée quand on emploie différentes valeurs de K ou (K, A). Pour maintenir cette additivité, on a fixé des valeurs communes de K = 0.4 et A = 0.4 dans le calcul des estimations de totaux et de changements. Les observations suivantes portent sur les résultats obtenus à partir de l'estimateur AK pour les valeurs K = 0.4 et A = 0.4. La dernière colonne du tableau 6 montre que la marge d'efficacité relative des estimations composites AK de totaux est de 6 à 10 % pour toutes les caractéristiques sauf le nombre de chômeurs. Les résultats présentés au tableau 7 révèlent que la marge d'efficacité relative des estimations composites AK de changements varie de 12 à 15 % dans tous les cas à l'exception de la caractéristique "nombre de chômeurs". Le gain d'efficacité de l'estimateur composite AK de totaux et de changements par rapport à l'estimateur simple par quotient est de 2 à 3 % pour la caractéristique "nombre de chômeurs".

REMERCIEMENTS

Les auteurs aimeraient remercier D. Drew, G.B. Gray et M.P. Singh de Statistique Canada pour leurs remarques pertinentes pendant la préparation de ce travail, ainsi que l'arbitre dont les suggestions ont contribué à améliorer la présentation de ce document.

TABLEAU 1

Correspondance entre les panels dans l'échantillon de l'EPA au mois m et les panels des mois m - j

Panels du mois m	m-1	m-2	m-3	m-4	m-5	m-6	m-7	m-8	m-9	m-10	m-11
1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))	((4))	((3))	((2))
2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))	((4))	((3))
3	2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))	((4))
4	3	2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))	((5))
5	4	3	2	1	(6)	(5)	(4)	(3)	(2)	(1)	((6))
6	5	4	3	2	1	(6)	(5)	(4)	(3)	(2)	(1)

Le coefficient de corrélation des panels qui font partie de l'échantillon aux mois m et m-j (les panels non entourés de parenthèses) est ρ_j.

Le coefficient de corrélation des panels du mois m et de leur premier prédecesseur (panel entouré d'une paire de parenthèses) au mois m - j est γ_j.

Le coefficient de corrélation des panels du mois m et de leur deuxième prédécesseur (panel entouré de deux paires de parenthèses) au mois m - j est $\gamma_j^{(2)}$. On suppose ici que tous les $\gamma_j^{(2)}$ sont nuls.

TABLEAU 2

Estimations des corrélations ρ (Ontario, 1980-1981)

Caractéristiques \ ρ	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	$\hat{\rho}_5$
Taille de la population active	.843	.782	.717	.674	.631
Nombre de personnes occupées	.852	.779	.709	.664	.619
Nombre de personnes occupées dans l'agriculture	.955	.926	.901	.861	.821
Nombre de personnes occupées dans les secteurs non agricoles	.861	.791	.724	.678	.632
Nombre de chômeurs	.580	.445	.334	.286	.238

TABLEAU 3

Estimations des corrélations γ (Ontario, 1980-1981)

Caractéristiques \ $\hat{\gamma}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$	$\hat{\gamma}_7$	$\hat{\gamma}_8$	$\hat{\gamma}_9$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{11}$
Taille de la population active	.161	.141	.128	.133	.135	.136	.125	.127	.124	.122	.127
Nombre de personnes occupées	.164	.136	.142	.142	.146	.149	.148	.150	.153	.141	.148
Nombre de personnes occupées dans l'agriculture	.477	.483	.474	.486	.480	.474	.459	.429	.394	.323	.252
Nombre de personnes occupées dans les secteurs non agricoles	.184	.150	.147	.157	.162	.167	.166	.169	.174	.156	.166
Nombre de chômeurs	.141	.074	.076	.063	.057	.051	.045	.060	.077	.136	.074

TABLEAU 4A & 4B
Valeurs optimales de (K, A) et K et indices d'efficacité relative des estimateurs composites K et AK

TABLEAU 4A
$\gamma_i \neq 0$

Caractéristiques	Estimateur composite K		Estimateur composite AK		
	Val. opt. K	Efficacité relative	Val. opt. K	A	Efficacité relative
Taille de la population active	0.7	118.8	0.8	0.48	128.4
Nombre de personnes occupées	0.7	118.5	0.8	0.49	128.1
Nombre de personnes occupées dans l'agriculture	0.8	120.6	0.8	0.38	126.9
Nombre de personnes occupées dans les secteurs non agricoles	0.7	119.4	0.8	0.47	129.3
Nombre de chômeurs	0.3	102.8	0.5	0.38	105.2

TABLEAU 4B

$\gamma_i = 0$ pour tous les i

Caractéristiques	Estimateur composite K		Estimateur composite AK		
	Val. opt. K	Efficacité relative	Val. opt. K	A	Efficacité relative
Taille de la population active	0.7	125.5	0.8	0.50	138.4
Nombre de personnes occupées	0.7	125.3	0.8	0.51	137.9
Nombre de personnes occupées dans l'agriculture	0.8	167.3	0.9	0.46	187.9
Nombre de personnes occupées dans les secteurs non agricoles	0.7	126.9	0.8	0.49	140.2
Nombre de chômeurs	0.4	104.4	0.6	0.51	108.4

L'efficacité relative compare la variance (V) de l'estimateur simple par quotient et celle de l'estimateur composite et se définit comme 100 fois le rapport V(estimateur simple)/V(estimateur K ou AK).

TABLEAU 5

Estimations (en milliers) du biais de renouvellement α_i

Caractéristiques	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_6$
Taille de la population active	-135.3	39.8	41.1	31.1	15.4	7.9
Nombre de personnes occupées	-141.7	35.5	34.9	31.3	25.4	14.8
Nombre de personnes occupées dans l'agriçulture	-4.2	-2.6	2.2	-0.1	4.2	0.5
Nombre de personnes occupées dans les secteurs non agricoles	-137.5	38.0	32.7	31.3	21.2	14.3
Nombre de chômeurs	6.4	4.3	6.2	-0.1	-9.9	-6.9

α_i est défini comme $E(y_{m,i}) - Y_m$ et est estimé par $y_{m,i} - \sum_{i=1}^{6} y_{m,i}/6$.

TABLEAU 6

Comparaison de la variance et de l'erreur quadratique moyenne de l'estimateur simple, des estimateurs composites K et AK

TAILLE DE LA POPULATION ACTIVE

	Estimateur simple	Estimateur composite K		Estimateur composite AK		Valeurs communes caractéristiques
		(EQM min.) K=0	(Var. min.) K=0.7	(EQM min.) K=0.7 A=0.7	(Var. min.) K=0.8 A=0.5	K=0.4 A=0.4
Estimation du total mensuel 10^3	4480.7	4480.7	4547.5	4484.4	4527.6	4481.7
Variance 10^6	432.0	432.0	363.8	358.1	336.5	391.8
Biais 10^3	0	0	66.8	3.7	46.9	1.0
Erreur quadratique moyenne 10^6	432.0	432.0	4284.5	371.5	2532.4	392.9
Efficacité relative		100.0	9.0	116.3	17.1	110.0

NOMBRE DE PERSONNES OCCUPÉES

		K=0	K=0.7	K=0.7 A=0.7	K=0.8 A=0.5	K=0.4 A=0.4
Estimation du total mensuel 10^3	4186.0	4186.0	4259.0	4183.6	4240.3	188.0
Variance 10^6	473.3	473.3	399.6	397.7	369.5	428.0
Biais 10^3	0	0	73.0	-2.4	54.3	2.0
Erreur quadratique moyenne 10^6	473.3	473.3	5732.2	403.2	3320.9	432.8
Efficacité relative		100.0	8.3	117.4	14.3	109.4

NOMBRE DE PERSONNES OCCUPÉES DANS L'AGRICULTURE (TABLEAU 6 suite)

	Estimateur simple	Estimateur composite K		Estimateur composite AK		Valeurs communes caractéristiques
		(EQM min.) K=0.6	(Var. min.) K=0.8	(EQM min.) K=0.8 A=0.6	(Var. min.) K=0.8 A=0.4	K=0.4 A=0.4
Estimation du total mensuel 10^3	142.0	143.4	145.7	143.2	144.1	142.1
Variance 10^6	85.7	75.6	71.1	68.7	67.6	80.8
Biais 10^3	0	1.4	3.7	1.2	2.1	0.1
Erreur quadratique moyenne 10^6	85.7	77.6	85.1	70.2	71.8	80.8
Efficacité relative		110.5	110.7	122.2	119.4	106.1

NOMBRE DE PERSONNES OCCUPÉES DANS LES SECTEURS NON AGRICOLES

		K=0	K=0.7	K=0.8 A=0.9	K=0.8 A=0.5	K=0.4 A=0.4
Estimation du total mensuel 10^3	4043.9	4043.9	4114.7	4041.6	4096.6	4045.8
Variance 10^6	498.9	498.9	417.8	418.0	385.9	452.8
Biais 10^3	0	0	70.8	-2.3	52.7	1.9
Erreur quadratique moyenne 10^6	498.9	498.9	5436.1	423.3	3161.7	456.4
Efficacité relative		100.0	9.2	117.9	15.8	109.3

NOMBRE DE CHÔMEURS (TABLEAU 6 fin)

	Estimateur simple	Estimateur composite K		Estimateur composite AK		Valeurs communes caractéristiques
		(EQM min.) K=0.2	(Var. min.) K=0.3	(EQM min.) K=0.4 A=0.4	(Var. min.) K=0.5 A=0.4	K=0.4 A=0.4
Estimation du total mensuel 10^3	294.8	294.1	293.7	293.9	293.2	293.9
Variance 10^6	117.5	114.9	414.3	112.5	111.7	112.5
Biais 10^3	0	-0.7	-1.1	-0.9	-1.6	-0.9
Erreur quadratique moyenne 10^6	117.5	115.4	115.7	113.3	114.4	113.3
Efficacité relative		101.9	101.6	103.7	102.7	103.7

Efficacité relative compare l'EMQ de l'estimateur simple par quotient et celle de l'estimateur composite et se définit comme 100 fois le rapport EQM(estimateur simple)/EQM(estimateur K ou AK).

TABLEAU 7

Efficacité relative des estimateurs composites du changement mensuel

Caractéristiques	Estimateur composite K		Estimateur composite AK			Valeurs communes K = 0.4 A = 0.4
	Val. opt. K	Efficacité relative	Val. opt. K	opt. A	Efficacité relative	Efficacité relative
Taille de la population active	0.9	146.6	0.9	0.1	147.9	113.3
Nombre de personnes occupées	0.9	151.0	0.9	0.1	152.3	114.1
Nombre de personnes occupées dans l'agriculture	0.9	234.7	0.9	0.0	234.7	112.3
Nombre de personnes occupées dans les secteurs non agricoles	0.9	154.0	0.9	0.1	155.2	114.1
Nombre de chômeurs	0.4	106.0	0.6	0.2	106.4	102.9

L'efficacité relative compare la variance (V) de l'estimateur simple par quotient et celle de l'estimateur composite et se définit comme 100 fois le rapport V(estimateur simple)/V(estimateur K ou AK).

ANNEXE I

Calcul du biais de l'estimateur composite

Tel qu'il est défini dans l'équation (2.2), l'estimateur composite AK de Y_m s'écrit

$$y'_m = (1 - K + A)y_{m,1}/6 + (1 - K - A/5) \sum_{j=2}^{6} y_{m,j}/6 + K(y'_{m-1} + d_{m,m-1}). \quad (A1.1)$$

Rappelons que l'estimateur simple actuellement utilisé dans l'EPA est la moyenne des estimateurs relatifs aux six panels définie par la formule :

$$\bar{y}_m = \sum_{i=1}^{6} y_{m,i}/6. \qquad (A1.2)$$

L'équation (2.3) précise que le biais de l'estimateur relatif au i^e panel est égal à α_i, de sorte que $E(y_{m,i}) = Y_m + \alpha_i$, où le biais est indépendant de m. Ainsi, on peut écrire

$E(\bar{y}_m) = Y_m + \sum_{i=1}^{6} \alpha_i/6 = Y_m + \bar{\alpha}$, et on suppose dans les raisonnements subséquents que $\bar{\alpha} = 0$.

L'estimateur composite peut se récrire de la manière suivante :

$$y'_m = y_m + K(y'_{m-1} + d_{m,m-1}), \qquad (A1.3)$$

où

$$y_m = (1 - K + A)y_{m,1}/6 + (1 - K - A/5)\sum_{j=2}^{6} y_{m,j}/6$$

$$= (1 - K)\bar{y}_m + A(y_{m,1} - \bar{y}_m)/5,$$

donc

$$E(y_m) = (1 - K)Y_m + \bar{\alpha}) + (A/5)(\alpha_1 - \bar{\alpha}), \qquad (A1.4)$$

qu'on peut simplifier sous la forme de l'équation

$$E(y_m) = (1 - K)Y_m + (A/5)\alpha_1, \text{ quand } \bar{\alpha} = 0.$$

Utilisant la définition de $d_{m,m-1}$, donnéee en (2.1), nous avons

$$E(d_{m,m-1}) = E[\sum_{j=2}^{6} (y_{m,j} - y_{m-1,j-1})/5]$$

$$= (Y_m - Y_{m-1}) + (\alpha_6 - \alpha_1)/5. \qquad (A1.5)$$

On peut développer y'_m par l'application récursive de l'équation (A1.3) et pour un décalage de n mois :

$$y'_m = y_m + Ky_{m-1} + K^2y_{m-2} + \ldots + K^{n-1}y_{m-n+1} + K^n y'_{m-n}$$

$$+ Kd_{m,m-1} + K^2 d_{m-1,\,m-2} + \ldots + K^n d_{m-n+1,\,m-n}. \qquad (A1.6)$$

L'espérance mathématique de y_m peut être facilement exprimée à partir des équations (A1.4) et (A1.6). Ainsi :

$$E(y'_m) = (1 - K)[Y_m + KY_{m-1} + K^2Y_{m-2} + \ldots + K^{n-1}Y_{m-n+1}] + K^n E(y'_{m-n})$$

$$+ [(1 - K)\,\bar{\alpha} + (A/5)(\alpha_1 - \bar{\alpha})](1 - K^n)/(1 - K)$$

$$+ K(Y_m - Y_{m-1}) + K^2(Y_{m-1} - Y_{m-2}) + \ldots + K^n(Y_{m-n+1} - Y_{m-n})$$

$$+ [(\alpha_6 - \alpha_1)/5]K(1 - K^n)/(1 - K)$$

$$= Y_m + K^n[E(y'_{m-n}) - Y_{m-n})]$$

$$+ [(1 - K)\bar{\alpha} + (A/5)(\alpha_1 - \bar{\alpha}) + K(\alpha_6 - \alpha_1)/5](1 - K^n)/(1 - K)$$

$$= Y_m + K^n[E(y'_{m-n}) - Y_{m-n}$$

$$+ [(1 - K - A/5)\bar{\alpha} + (A/5)\alpha_1 + K(\alpha_6 - \alpha_1)/5](1 - K^n)/(1 - K) \qquad (A1.7)$$

qu'on peut simplifier de la manière suivante pour une valeur suffisamment grande de m et le cas où $\bar{\alpha} = 0$:

$$E(y'_m) = Y_m + [A\alpha_1 + K(\alpha_6 - \alpha_1)]/[5(1 - K)]. \qquad (A1.8)$$

Comme le biais de y'_m dans le modèle construit ici est indépendant de m,

l'écart entre les estimations composites relatives à deux mois séparés par un intervalle r sont sans biais. Bref :

$$E(y'_m - y_{m-r}) = Y_m - Y_{m-r} \text{ pour tout } r. \qquad (A1.9)$$

ANNEXE II

Calcul de la variance de l'estimateur composite

Nous supposons que les estimateurs composites (voir l'équation (2.2)) sont devenus assez stables d'une période à l'autre et que, par conséquent, $V(y'_{m-1}) = V(y'_m)$. Comme toutes les corrélations à douze mois ou plus d'intervalle sont nulles par hypothèse, nous supposons que les estimateurs composites de l'EPA se sont stabilisés après douze mois.

En prenant la variance des deux membres de (A1.3) et appliquant l'hypothèse susmentionnée, on peut calculer $V(y'_m)$ et obtenir le résultat suivant :

$$V(y'_m) = [V(y_m) + K^2V(d_{m,m-1}) + 2KCov(y_m, d_{m,m-1}) + 2KCov(y_m, y'_{m-1}) + 2K^2Cov(d_{m,m-1}, y'_{m-1})]/(1 - K^2). \qquad (A2.1)$$

Pour éliminer y'_{m-1} du membre de droit de (A2.1), on peut prendre l'équation (A1.6), remplacer m par (m - 1) et n par 12 pour obtenir :

$$y'_{m-1} = \sum_{g=1}^{12} (K^{g-1}y'_{m-g} + K^g d_{m-g,m-g-1}) + K^{12}y'_{m-13}. \qquad (A2.2)$$

Si on reporte l'équation (A2.2) dans l'équation (A2.1) et qu'on supprime les termes nuls, le résultat est :

$$V(y'_m) = [V(y_m) + K^2V(d_{m,m-1}) + 2KCov(y_m, d_{m,m-1})$$

$$+ 2 \sum_{g=1}^{12} K^g[\mathrm{Cov}(y_m, y_{m-g}) + K\mathrm{Cov}(d_{m,m-1}, y_{m-g})$$

$$+ K\mathrm{Cov}(y_m, d_{m-g,m-g-1})$$

$$+ K^2\mathrm{Cov}(d_{m,m-1}, d_{m-g,m-g-1})]\}/(1 - K^2). \qquad (A2.3)$$

Les expressions qui définissent les variances et les covariances qui figurent dans le membre de droite de (A2.3) peuvent être facilement calculées à partir des équations (2.1) et (A1.3). Ainsi, on obtient

$$V(y_m) = [(1 - K^2/6 + A^2/30]\sigma^2, \qquad (A2.4)$$

qui correspond simplement à $\sigma^2/6$ quand A = K = 0; par conséquent,

$$V(\bar{y}_m) = \sigma^2/6 \qquad (A2.4a)$$

est la variance de l'estimateur utilisé actuellement dans l'EPA.

$$V(d_{m,m-1}) = 2\sigma^2(1 - \rho_1)/5. \qquad (A2.5)$$

$$\mathrm{Cov}(y_m, d_{m,m-1}) = (1 - K)(1 - \rho_1)\sigma^2/6 - A(1 - \rho_1)\sigma^2/30. \qquad (A2.6)$$

Pour calculer les autres covariances de l'équation (A2.3) qui comportent un retard g, définissons une fonction indicatrice I(a, b) :

$$I(a, b) = 1 \text{ si } a \leq b$$
$$= 0 \text{ autrement.}$$

Si on examine la définition de y_m dans l'équation (A1.3) et de $d_{m,m-1}$ dans l'équation (2.1) et les expressions correspondantes pour le mois (m - g), on peut conclure que les covariances suivantes sont nécessaires pour calculer les autres covariances de (A2.3) :

$$\mathrm{Cov}(\bar{y}_m, \bar{y}_{m-g}) = (\sigma^2/36)[(6 - g)\rho_g I(g, 5) + (6 - |\gamma - 6|)\gamma_g],$$

$$\mathrm{Cov}(\bar{y}_m,\ y_{m-q-1}) = (\sigma^2/6)[\rho_q I(q,\ 5) + \gamma_q I\ (6,\ q)],$$

$$\mathrm{Cov}(y_{m,1},\ \bar{y}_{m-q}) = (\sigma^2/6)\gamma_q I(q,\ 6),$$

$$\mathrm{Cov}(y_{m,1},\ y_{m-q,1}) = \sigma^2 I(q,\ 6)I(6,\ q)\gamma_6 \text{ (expression nulle sauf quand } q = 6).$$

$$\mathrm{Cov}(y_{m-1,6},\ \bar{y}_{m-q}) = (\sigma^2/6)[\rho_{q-1} I(q,\ 6) + \gamma_{q-1} I(7,\ q)],$$

$$\mathrm{Cov}(y_{m-1,6},\ y_{m-q-1}) = \sigma^2[\rho_{q-1} I(q,\ 6)I(6,\ q)],$$

$$\mathrm{Cov}(\bar{y}_m,\ y_{m-q-1,6}) = (\sigma^2/6)\gamma_{q+1} I(q,\ 5),$$

$$\mathrm{Cov}(y_{m,1},\ y_{m-q-1,6}) = \sigma^2\gamma_1 I(q,\ 0)I(0,\ q) \text{ (expression nulle pour } q \geq 1) \quad \text{(A2.7)}$$

Les quatre covariances de (A2.3) qui comprennent le retard q peuvent alors être définies comme suit :

$$\begin{aligned}\mathrm{Cov}(y_m,\ y_{m-q}) = {} & \sigma^2\rho_q I(q,\ 5)[(1 - K)^2(6 - q)/36 \\ & + A(1 - K)(q - 3)/90 - qA^2/900] \\ & + \sigma^2\gamma_q[(1 - K)^2(6 - |q - 6|)/36 \\ & + A(1 - K)(|q - 6| - 3)/90 - |q - 6|A^2/900] \\ & + \sigma^2\gamma_q I(q,\ 6)I(6,\ q)A(1 - K + A)/30. \end{aligned} \quad \text{(A2.8)}$$

$$\begin{aligned}\mathrm{Cov}(d_{m,m-1},\ y_{m-q}) = {} & \sigma^2(\rho_q - \rho_{q-1})I(q,\ 5)[(1 - K)(6 - q)/30 + qA/150] \\ & + \sigma^2(\gamma_q - \gamma_{q-1})[(1 - K)(6 - |q - 6|)/30 + |q - 6|A/150 \\ & - (1 - K + A)I(q,\ 6)/30]. \end{aligned} \quad \text{(A2.9)}$$

$$\mathrm{Cov}(y_m,\ d_{m-q,m-q-1}) = \sigma^2(\rho_q - \rho_{q+1})I(q,\ 5)(1 - K - A/5)(5 - q)/30,$$

$$+ \sigma^2(\gamma_q - \gamma_{q+1})[(1 - K - A/5)(6 - I(6, q) - |q - 6|)/30 + AI(q, 5)/25], \quad (A2.10)$$

$$\mathrm{Cov}(d_{m,m-1}, d_{m-q, m-q-1}) = \sigma^2[(5 - q)(2\rho_q - \rho_{q-1} - \rho_{q+1})I(q, 5) + 5 - |q - 6|)(2\gamma_q - \gamma_{q-1} - \gamma_{q+1})]/25. \quad (A2.11)$$

On peut donc exprimer $V(y'_m)$ sous la forme d'une équation de type $aA^2 + bA + c = f(A)$, où a, b et c sont des fonctions de K, des ρ et des γ. On peut démontrer que $a \geq 0$. La valeur de A qui réduit au minimum la variance de l'estimateur AK a été déterminé pour K = 0(0.1)0.9. Parmi toute la série de couples (A, K) dans cet intervalle, la valeur optimale de (A, K) a été choisie et c'est ce couple qui est présenté au tableau 4A.

ANNEXE III

Calcul de la variance de $Y'_m - Y'_{m-1}$

Prenons l'équation (A1.3) :

$y'_m = y_m + K(y'_{m-1} + d_{m,m-1})$, qu'on peut récrire de la manière suivante :

$y'_m - Ky'_{m-1} = y_m + Kd_{m,m-1}$,

d'où il découle que

$$(1 + K^2)V(y'_m) - 2K\mathrm{Cov}(y'_m, y'_{m-1}) = V(y_m) + 2K\mathrm{Cov}(y_m, d_{m,m-1}) + K^2V(d_{m,m-1}).$$

Quand $K \neq 0$, on peut obtenir l'expression de $\mathrm{Cov}(y'_m, y'_{m-1})$ à partir de l'équation ci-dessus et, en faisant des substitutions à l'aide des équations (A2.4),

(A2.5) et (A2.6) et en utilisant le fait que $V(y'_m - y'_{m-1}) = 2V(y'_m) - 2Cov(y'_m, y'_{m-1})$, on obtient le résultat suivant pour $K \neq 0$:

$$V(y'_m - y'_{m-1}) = \sigma^2[A^2/30 - (1 - \rho_1)KA/15 + (1 - K)^2/6 + (1 - \rho_1)K(K + 5)/15]/K - (1 - K)^2V(y'_m)/K. \quad (A3.1)$$

Quand $K = 0$,

$$y'_m = (1 - A/5)\bar{y}_m + Ay_{m,1}/5, \text{ et}$$

$$Cov(y'_m, y'_{m-1}) = Cov[(1 - A/5)\bar{y}_m + Ay_{m,1}/5, (1 - A/5)\bar{y}_{m-1} + Ay_{m-1,1}/5]. \quad (A3.2)$$

Ainsi, pour $K = 0$:

$$V(y'_m - y'_{m-1}) = \sigma^2[(1/15 + \rho_1/450 + \gamma_1/90)A^2 + 2(\rho_1 - \gamma_1)A/45 + 1/3 - (5\rho_1 + \gamma_1)/18]. \quad (A3.3)$$

BIBLIOGRAPHIE

[1] Bailar, B.A., "The Effects of Rotation Group Bias on Estimates from Panel Surveys", Journal of the American Statistical Association, 70, 23-29, 1975.

[2] Ghangurde, P.D., "Le biais de renouvellement de l'échantillon dans les estimations de l'EPA", Techniques d'enquête, 8, 86-101, 1982.

[3] Gurney, M. et Daly, J.F., "A Multivariate Approach to Estimation in Periodic Sample Surveys", Proceedings of the Social Statistics Section, American Statistical Association, 242-257, 1965.